INSIGHTS FIFA 2019 DATA

FELIPE

20/12/2020

ANÁLISE EXPLORATÓRIA DA BASE DE DADOS FIFA 2019

Bibliotecas usadas.

```
library(tidyverse)
library(data.table)
library(readxl)
library(na.tools)
library(magrittr)
library(dplyr)
library(lubridate)
library(stringr)
```

Baixando base de dados.

```
base_fifa_2019<-read_csv("data.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   X1 = col_double(),
##   ID = col_double(),
##   Age = col_double(),
##   Overall = col_double(),
##   Potential = col_double(),
##   Special = col_double(),
##   `International Reputation` = col_double(),
##   `Weak Foot` = col_double(),
##   `Skill Moves` = col_double(),
##   `Jersey Number` = col_double(),
##   Crossing = col_double(),
##   Finishing = col_double(),
##   HeadingAccuracy = col_double(),
##   ShortPassing = col_double(),
```

```
##   Volleys = col_double(),
##   Dribbling = col_double(),
##   Curve = col_double(),
##   FKAccuracy = col_double(),
##   LongPassing = col_double(),
##   BallControl = col_double()
##   # ... with 24 more columns
## )
## i Use `spec()` for the full column specifications.
```

Excluindo primeira variável.

```
base_fifa_2019<-base_fifa_2019[,-1]
```

Exibindo base de dados.

```
base_fifa_2019
```

```
## # A tibble: 18,207 x 88
##        ID Name    Age Photo Nationality Flag  Overall Potential Club
##     <dbl> <chr> <dbl> <chr> <chr>       <chr>   <dbl>     <dbl> <chr>
##  1 158023 L. M~    31 http~ Argentina   http~      94        94 FC B~
##  2  20801 Cris~    33 http~ Portugal    http~      94        94 Juve~
##  3 190871 Neym~    26 http~ Brazil      http~      92        93 Pari~
##  4 193080 De G~    27 http~ Spain       http~      91        93 Manc~
##  5 192985 K. D~    27 http~ Belgium     http~      91        92 Manc~
##  6 183277 E. H~    27 http~ Belgium     http~      91        91 Chel~
##  7 177003 L. M~    32 http~ Croatia     http~      91        91 Real~
##  8 176580 L. S~    31 http~ Uruguay     http~      91        91 FC B~
##  9 155862 Serg~    32 http~ Spain       http~      91        91 Real~
## 10 200389 J. O~    25 http~ Slovenia    http~      90        93 Atlé~
## # ... with 18,197 more rows, and 79 more variables: `Club Logo` <chr>,
## #   Value <chr>, Wage <chr>, Special <dbl>, `Preferred Foot` <chr>,
## #   `International Reputation` <dbl>, `Weak Foot` <dbl>, `Skill Moves` <dbl>,
## #   `Work Rate` <chr>, `Body Type` <chr>, `Real Face` <chr>, Position <chr>,
## #   `Jersey Number` <dbl>, Joined <chr>, `Loaned From` <chr>, `Contract Valid
## #   Until` <chr>, Height <chr>, Weight <chr>, LS <chr>, ST <chr>, RS <chr>,
## #   LW <chr>, LF <chr>, CF <chr>, RF <chr>, RW <chr>, LAM <chr>, CAM <chr>,
## #   RAM <chr>, LM <chr>, LCM <chr>, CM <chr>, RCM <chr>, RM <chr>, LWB <chr>,
## #   LDM <chr>, CDM <chr>, RDM <chr>, RWB <chr>, LB <chr>, LCB <chr>, CB <chr>,
## #   RCB <chr>, RB <chr>, Crossing <dbl>, Finishing <dbl>,
## #   HeadingAccuracy <dbl>, ShortPassing <dbl>, Volleys <dbl>, Dribbling <dbl>,
## #   Curve <dbl>, FKAccuracy <dbl>, LongPassing <dbl>, BallControl <dbl>,
## #   Acceleration <dbl>, SprintSpeed <dbl>, Agility <dbl>, Reactions <dbl>,
## #   Balance <dbl>, ShotPower <dbl>, Jumping <dbl>, Stamina <dbl>,
## #   Strength <dbl>, LongShots <dbl>, Aggression <dbl>, Interceptions <dbl>,
## #   Positioning <dbl>, Vision <dbl>, Penalties <dbl>, Composure <dbl>,
## #   Marking <dbl>, StandingTackle <dbl>, SlidingTackle <dbl>, GKDiving <dbl>,
## #   GKHandling <dbl>, GKKicking <dbl>, GKPositioning <dbl>, GKReflexes <dbl>,
## #   `Release Clause` <chr>
```

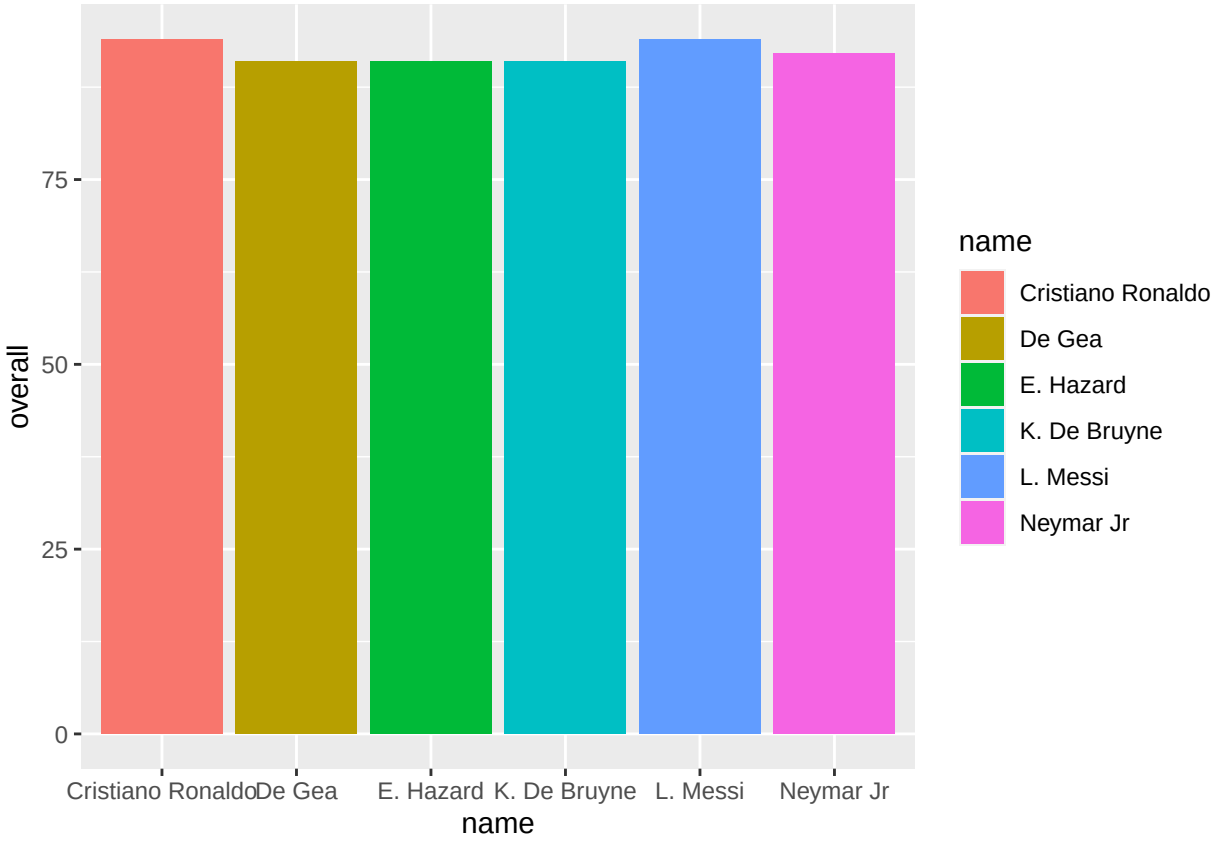
NOMES VARIAVEIS - minúsculas.

```
names(base_fifa_2019)<-str_to_lower(string = names(base_fifa_2019),locale = "en")
names(base_fifa_2019)
```

```
##  [1] "id"                         "name"
##  [3] "age"                        "photo"
##  [5] "nationality"                "flag"
##  [7] "overall"                    "potential"
##  [9] "club"                       "club logo"
## [11] "value"                      "wage"
## [13] "special"                    "preferred foot"
## [15] "international reputation"   "weak foot"
## [17] "skill moves"                "work rate"
## [19] "body type"                  "real face"
## [21] "position"                   "jersey number"
## [23] "joined"                     "loaned from"
## [25] "contract valid until"       "height"
## [27] "weight"                     "ls"
## [29] "st"                         "rs"
## [31] "lw"                         "lf"
## [33] "cf"                         "rf"
## [35] "rw"                         "lam"
## [37] "cam"                        "ram"
## [39] "lm"                         "lcm"
## [41] "cm"                         "rcm"
## [43] "rm"                         "lwb"
## [45] "ldm"                        "cdm"
## [47] "rdm"                        "rwb"
## [49] "lb"                         "lcb"
## [51] "cb"                         "rcb"
## [53] "rb"                         "crossing"
## [55] "finishing"                  "headingaccuracy"
## [57] "shortpassing"               "volleys"
## [59] "dribbling"                  "curve"
## [61] "fkaccuracy"                 "longpassing"
## [63] "ballcontrol"                "acceleration"
## [65] "sprintspeed"                "agility"
## [67] "reactions"                  "balance"
## [69] "shotpower"                  "jumping"
## [71] "stamina"                    "strength"
## [73] "longshots"                  "aggression"
## [75] "interceptions"              "positioning"
## [77] "vision"                     "penalties"
## [79] "composure"                  "marking"
## [81] "standingtackle"             "slidingtackle"
## [83] "gkdiving"                   "gkhandling"
## [85] "gkkicking"                  "gkpositioning"
## [87] "gkreflexes"                 "release clause"
```

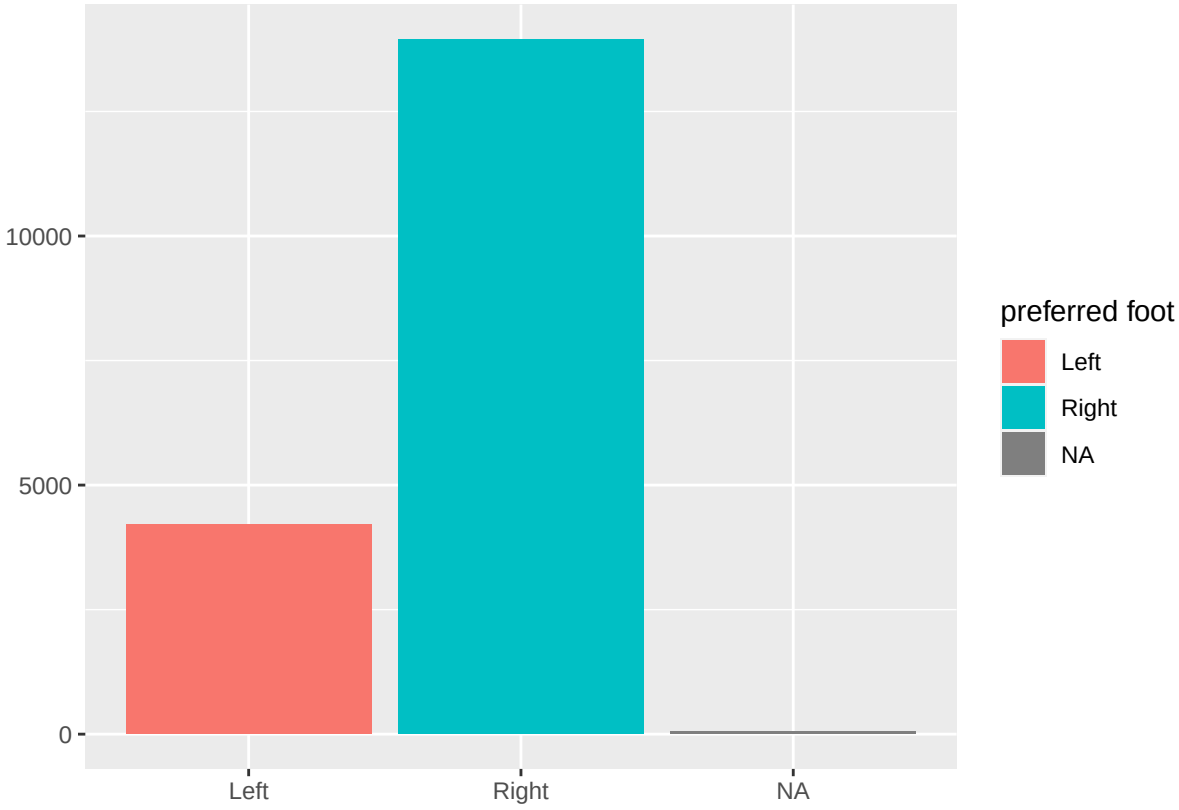
Melhores Jogadores - OVERALL.

```
base_fifa_2019 %>% arrange(desc(overall)) %>% head() %>%
  ggplot()+
  geom_col(mapping = aes(x = name,y =overall, fill = name))
```



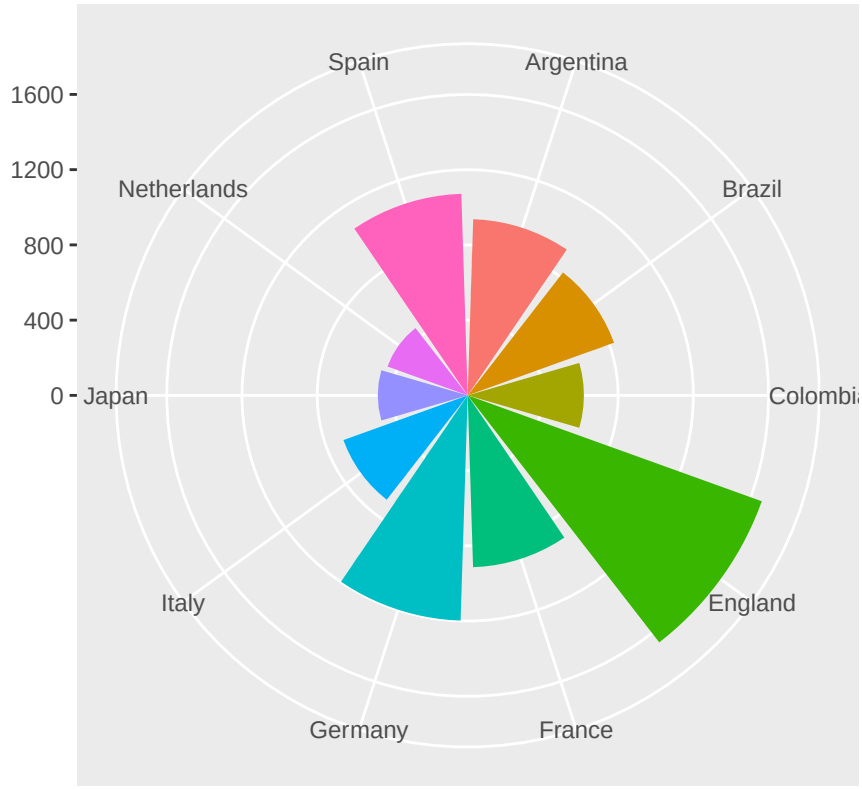
Números total de jogadores que jogam com o pé direito.

```
base_fifa_2019 %>%
  ggplot()+
    geom_bar(mapping = aes(x =  `preferred foot`, fill = `preferred foot`))+
   xlab(label = " ")+
    ylab(label = " ")
```



Distribuição das nacionalidades dos jogadores .. 10 primeiras posições.

```
base_fifa_2019 %>% count(nationality) %>% arrange(desc(n)) %>% head(10) %>%
  ggplot()+
    geom_col(mapping = aes(x = nationality,y = n , fill = nationality ), show.legend = FALSE) +
      coord_polar()+
         xlab(label = " ")+
          ylab(label = " ")
```



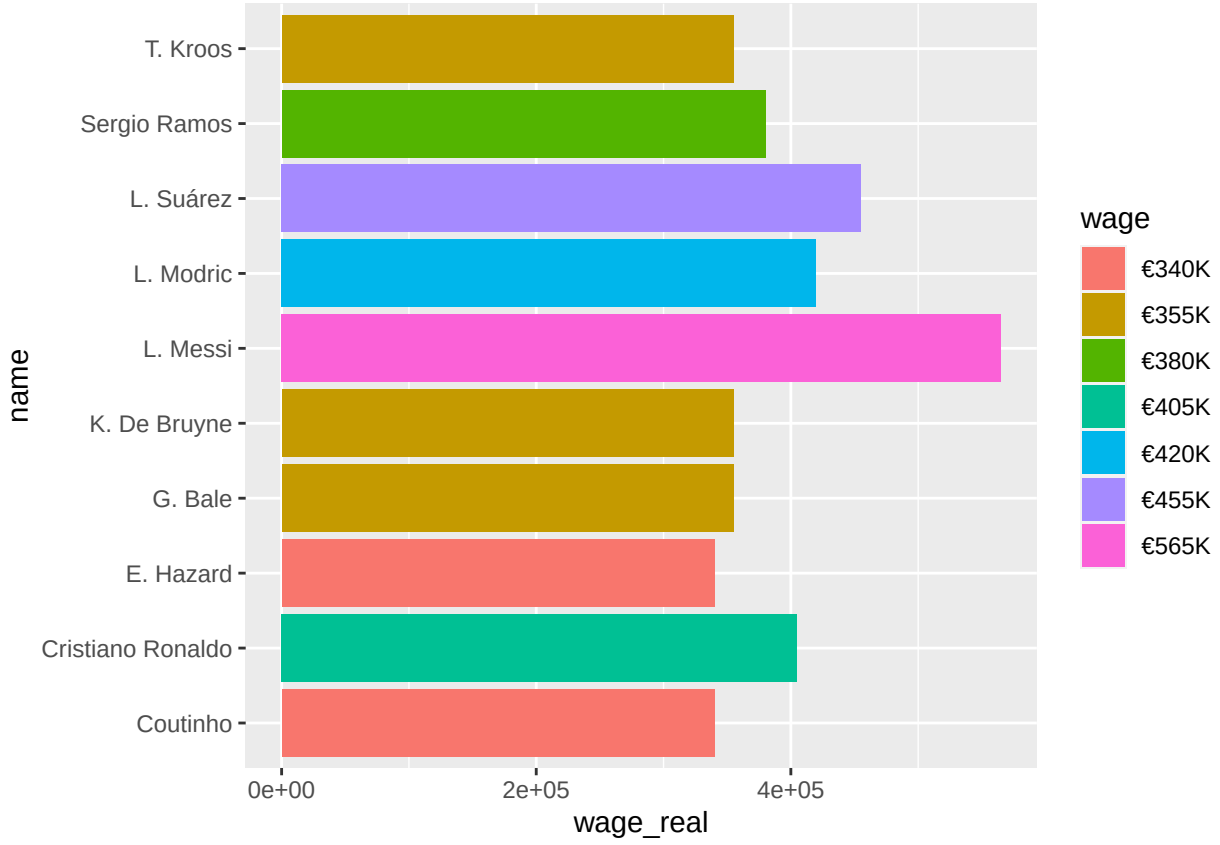
Maiores sálarios... 10 primeiros.

```
wage_Real<-NULL

for (i in 1:length(base_fifa_2019$wage)) {

  if (str_length(base_fifa_2019$wage[i]) == 5) {
    wage_Real[i]<-str_sub(string = base_fifa_2019$wage[i],start = 2,end = 4)
  }
  if (str_length(base_fifa_2019$wage[i]) == 4) {
    wage_Real[i]<-str_sub(string = base_fifa_2019$wage[i],start = 2,end = 3)
  }
  if (str_length(base_fifa_2019$wage[i]) == 3) {
    wage_Real[i]<-str_sub(string = base_fifa_2019$wage[i],start = 2,end = 2)
  }
  if (str_length(base_fifa_2019$wage[i]) == 2) {
    wage_Real[i]<-str_sub(string = base_fifa_2019$wage[i],start = 2,end = 2)
  }
}

base_fifa_2019 %>% select(name) %>% mutate(wage_real=as.double(wage_Real) * 1000) %>% arrange(desc(wage_
```

```
## # A tibble: 18,207 x 2
##    name              wage_real
```

```
##    <chr>                <dbl>
##  1 L. Messi            565000
##  2 L. Suárez           455000
##  3 L. Modric           420000
##  4 Cristiano Ronaldo   405000
##  5 Sergio Ramos        380000
##  6 K. De Bruyne        355000
##  7 T. Kroos            355000
##  8 G. Bale             355000
##  9 E. Hazard           340000
## 10 Coutinho            340000
## # ... with 18,197 more rows
```

```
base_fifa_2019 %>% select(name,wage) %>% mutate(wage_real=as.double(wage_Real) * 1000) %>% arrange(desc
  head(10) %>%
    ggplot(mapping = aes(x = name,y = wage_real,fill = wage))+
      geom_col()+
    coord_flip()
```



Avaliação média de jogadores por país.

```
base_fifa_2019 %>% group_by(nationality) %>% summarize(Average_Rating=mean(overall,na.rm = TRUE),quanti
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 164 x 3
##    nationality          Average_Rating quantity
##    <chr>                         <dbl>    <int>
##  1 United Arab Emirates           77          1
##  2 Central African Rep.           73.3        3
##  3 Israel                         72.1       14
##  4 Dominican Republic             72          2
##  5 Oman                           72          1
##  6 São Tomé & Príncipe            72          1
##  7 Cape Verde                     71.6       19
##  8 Portugal                       71.3      322
##  9 Togo                           71.3       12
## 10 Brazil                         71.3      827
## # ... with 154 more rows
```

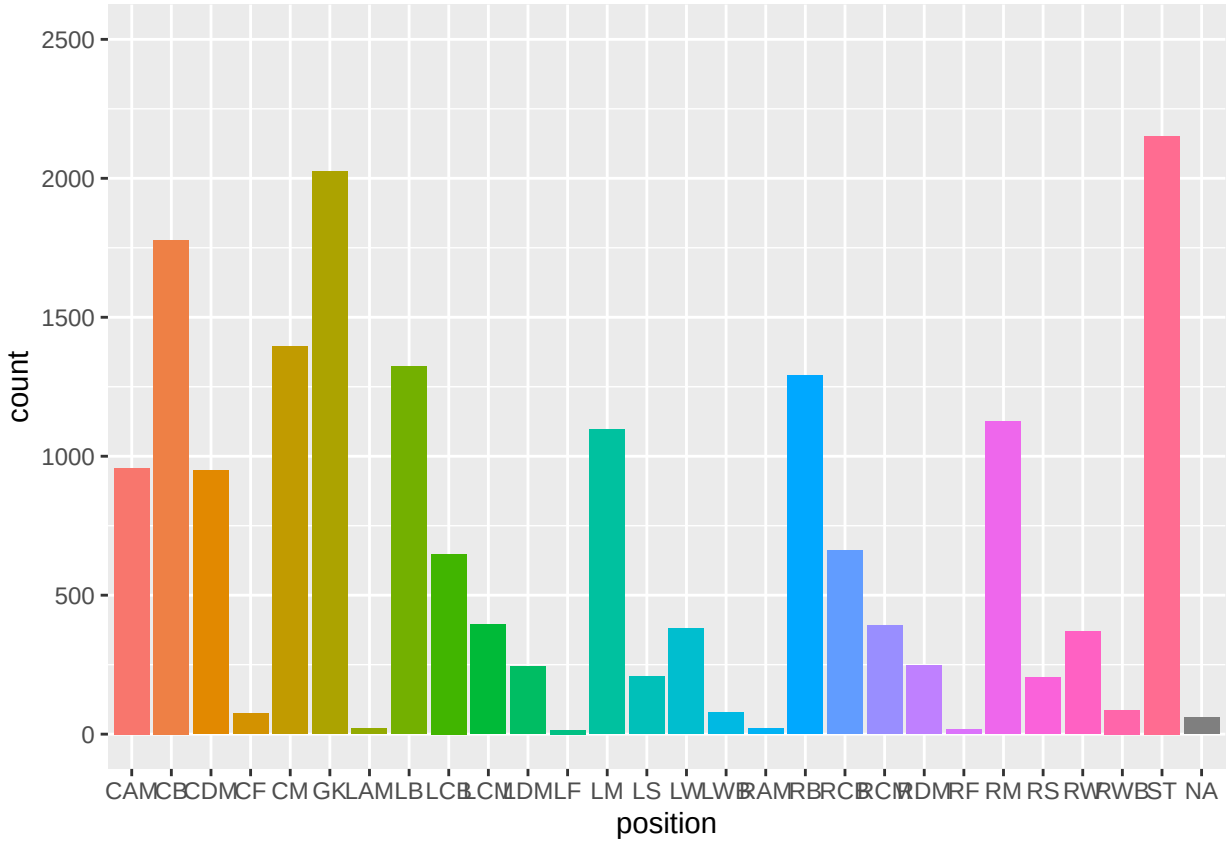
Distribuiçao de jogadores por posição.

```
base_fifa_2019 %>% count(position) %>% arrange(desc(n))
```

```
## # A tibble: 28 x 2
##    position     n
##    <chr>    <int>
##  1 ST        2152
##  2 GK        2025
##  3 CB        1778
##  4 CM        1394
##  5 LB        1322
##  6 RB        1291
##  7 RM        1124
##  8 LM        1095
##  9 CAM        958
## 10 CDM        948
## # ... with 18 more rows
```

Grafico de distribuiçao de jogadores por posição.

```
 base_fifa_2019 %>%
  ggplot()+
geom_bar(mapping = aes(x = position,fill = position),show.legend = FALSE)+
 ylim(0,2500)
```



Quando o contrato irá expirar.

```
base_fifa_2019 %>% select(name,`contract valid until`)
```

```
## # A tibble: 18,207 x 2
##    name              `contract valid until`
##    <chr>             <chr>
##  1 L. Messi          2021
##  2 Cristiano Ronaldo 2022
##  3 Neymar Jr         2022
##  4 De Gea            2020
##  5 K. De Bruyne      2023
##  6 E. Hazard         2020
##  7 L. Modric         2020
##  8 L. Suárez         2021
##  9 Sergio Ramos      2020
## 10 J. Oblak          2021
## # ... with 18,197 more rows
```